УДК 519.876:55.435.62(477.75)

***В.М. Таран*[1], *А.О. Пашко*[2]**
[1]Європейський університет, Ялтинська філія, м. Ялта, Україна
victoriya_yalta@ukr.net
[2]Європейський університет, м. Київ, Україна
pashkoua@mail.ru

Оцінювання прогнозів зсувних процесів Південного берега Криму за допомогою аналізу часових рядів (АРПКС)

Досліджено загальні закономірності побудови авторегресійних моделей зсувного процесу на прикладі Південного берега Криму. Проведено ідентифікацію АРПКС-моделей та розрахунки оцінок параметрів цих моделей. Обрано критерії оцінювання моделей та оцінювання прогнозів, за якими виділяється модель серед інших, та побудовано прогноз на 10 років.

Вступ

Задача аналізу, моделювання та прогнозування зсувних процесів Південного берега Криму обумовлена присутністю в даній системі чинника ризику непередбачених ситуацій та ускладнена невизначеністю зміни кліматичних, екологічних, геологічних та техногенних процесів, які негативно впливають на господарські об’єкти та інженерно-будівельні конструкції, на погіршення їх надійності і довговічності, а також призводять до їх руйнації [1].

Таким чином, стає актуальною проблема системного аналізу динаміки зсувних процесів, а також оцінювання прогнозів щодо їх активізації, можливих збитків та обсягу коштів, які знадобляться для покриття матеріальних збитків від руйнації об’єктів господарювання внаслідок зсувів.

Отже, стає необхідною розробка нових моделей, методик, алгоритмів, інтелектуальних систем, які б допомагали особі, що приймає рішення, у виробленні і прийнятті управлінських рішень в умовах невизначеності.

У наукових дослідженнях, що присвячені прогнозуванню зсувних процесів Південного берега Криму, слабкою ланкою є спрямованість лише на спостереження, моніторинг, картування та експертну оцінку фахівця, відсутність системного підходу, обмежене використання сучасних інформаційних технологій та інтелектуальних систем прийняття рішень.

Метою даної роботи є розвиток методики прогнозування зсувних процесів Південного берега Криму з використанням регресійного аналізу, задля чого виконується дослідження і аналіз факторів. Основним критерієм при побудові моделей обрано адекватність спостереженням, точність прогнозу та стійкість моделі.

Визначена мета статті зумовила необхідність вирішення таких задач:

- аналіз факторів та змінних, які описують зсувні процеси;
- розробка математичних моделей – авторегресії, АРПКС, їх аналіз та оцінку;
- побудова прогнозу зсувних процесів Південного берега Криму.

У рамках даної роботи розглядається задача оцінювання оперативного прогнозу зсувних процесів Південного берега Криму, що є складовою задачі управління екзогенними процесами і пов’язана з мінімізацією вимушених витрат на відновлення ушкоджених об’єктів. Вона має складний характер внаслідок недостатньої формалізації та структурованості, залежності від поточної ситуації, високого рівня невизначеності, обмеженості у часі для прийняття рішення.

Методика Бокса-Дженкінса для побудови математичної моделі

Методика побудови математичної моделі процесу у вигляді різницевого рівняння складається з наступних кроків [2]:

– обчислення та аналіз автокореляційної та часткової автокореляційної функцій (АКФ та ЧАКФ);

– оцінювання коефіцієнтів математичної моделі вибраної структури (чи багатьох структур) за допомогою вибраного методу оцінювання (найчастіше це МНК або його модифікації);

– діагностика отриманих моделей та вибір з них моделі, адекватної процесу, що моделюється.

Розглянемо детальніше згадані етапи побудови моделі.

На першому етапі виконують нормування та візуальну перевірку експериментальних (статистичних) даних і за необхідності коригують їх. Коригування даних полягає у заповненні пропусків та зменшенні викидів, що виходять за діапазон допустимих значень змінних. Нормування даних означає їх логарифмування або приведення до зручного діапазону їх зміни, наприклад, від 0 до 1, або від –1 до +1. Дуже часто для побудови моделі використовують не самі абсолютні значення статистичних даних, а їх перші чи другі різниці. Поширеним методом нормування даних є їх логарифмування з наступним формуванням додаткових часових рядів з перших чи других різниць. Як правило, із значень ряду віднімається його середнє значення, щоб отримати можливість працювати з відхиленнями, а не абсолютними значеннями змінних. Застосування того чи іншого методу для нормування даних може визначатися в кожному випадку по-своєму.

АКФ та ЧАКФ використовують для визначення попередньої оцінки порядку моделі, тобто скільки затриманих в часі значень необхідно брати для описання процесу. При цьому необхідно врахувати, що АКФ дає менш «чітку» оцінку порядку процесу. Наприклад, для процесу АР(1) значення основної змінної $y(k)$ та $y(k-2)$ будуть корельованими, не дивлячись на те, що $y(k-2)$ не присутнє в моделі. Кореляція між $y(k)$ і $y(k-2)$, тобто ρ_2, дорівнює коефіцієнту кореляції між значеннями $y(k)$ і $y(k-1)$, помноженому на коефіцієнт кореляції між $y(k-1)$ і $y(k-2)$ або $\rho_2 = \rho_1\rho_1 = \rho_1^2$. Подібні «непрямі» кореляції присутні в АКФ будь-якого процесу авторегресії. Нагадаємо, що вибіркова АКФ обчислюється за виразом:

$$r_S = \frac{\sum_{k=s+1}^{N}[y(k)-\mu][y(k-s)-\mu]}{\sum_{k=1}^{N}[y(k)-\mu]^2},$$

де N – кількість значень у вибірці даних; μ – середнє значення ряду.

Для стаціонарного процесу (це процес із постійними середнім значенням, дисперсією та коваріацією) коефіцієнти r_s мають нормальне розподілення та нульове середнє.

На відміну від АКФ, часткова АКФ між значеннями $y(k)$ та $y(k-s)$ виключає вплив величин $y(k-1)$ … $y(k-s)$, а це означає, що коефіцієнти ЧАКФ більш чітко відображають зв'язок між окремими значеннями основної змінної. Так, для процесу АР(1) ЧАКФ між $y(k)$ та $y(k-2)$ дорівнює нулю за визначенням, що підтверджується обчисленими значеннями ЧАКФ. Для того, щоб знайти попередню оцінку порядку моделі, вибіркові коефіцієнти ЧАКФ (тобто коефіцієнти, знайдені за вибіркою даних) можна обчислити за допомогою простого методу, який полягає в наступному.

а) Формують додатковий часовий ряд із відхилень основної змінної:

$$\{y'(k)\} = \{y(k)\} - \mu,$$

де μ – середнє значення ряду.

б) Формують рівняння першого порядку

$$y'(k) = \Phi_{11}\, y'(k-1) + e(k),$$

де $e(k)$ – похибка моделі. В такому рівнянні Φ_{11} відіграє роль коефіцієнта АКФ та ЧАКФ між $y(k)$ та $y(k-1)$. Для оцінювання двох коефіцієнтів можна сформувати рівняння другого порядку:

$$y'(k) = \Phi_{11}\, y'(k-1) + \Phi_{22}\, y'(k-2) + e(k),$$

де Φ_{22} – коефіцієнт ЧАКФ між $y(k)$ та $y(k-2)$.

У загальному випадку коефіцієнти ЧАКФ стаціонарного процесу АРСС(p,q) повинні збігатися до нуля, починаючи з p-го значення. АКФ процесу АРСС(p,q) починає збігатися до нуля при значеннях зміщення $s \geq q$.

На другому етапі оцінюють коефіцієнти (параметри) різницевого рівняння, використовуючи принцип економії або збереження. Цей принцип означає, що кількість коефіцієнтів, що оцінюються, не повинна перевищувати їх необхідне число.

При моделюванні процесів необхідно пам'ятати, що поведінку процесу необхідно апроксимувати за допомогою рівнянь, а не намагатися описати її до найменших дрібниць. Необхідно враховувати також, що різні за формою моделі можуть мати однакові властивості.

На третьому етапі отримана модель діагностується, тобто виконується перевірка на адекватність. Діагностика складається з наступних кроків.

а) Візуальне дослідження графіка похибок моделі $e(k) = y^*(k) - y(k)$, де $y^*(k)$ – оцінка змінної, отримана за допомогою рівняння. На графіку не повинно бути викидів та довгих інтервалів, на яких похибка приймає великі значення (тобто довгих інтервалів значної неадекватності).

б) Похибки моделі не повинні бути корельовані між собою. Для аналізу наявності кореляції між значеннями похибок необхідно обчислити АФ та ЧАКФ для ряду $\{e(k)\}$ і за допомогою Q-статистики визначити ступінь корельованості (наприклад, Q-статистика вважається несуттєвою до рівня 10%).

в) Для моделі 2 – 3 порядку оцінки параметрів повинні збігатися до усталених значень після 40 – 60 ітерацій алгоритму оцінювання. Якщо кількість ітерацій набагато перевищує вказані числа, то це свідчить про те, що процес може бути нестаціонарним.

г) Сума квадратів похибок повинна бути мінімальною у порівнянні з усіма іншими моделями, тобто:

$$\sum_{k=1}^{N} e^2(k) = \sum_{k=1}^{N} [y^*(k) - y(k)]^2 \to \min.$$

д) Для оцінки адекватності моделі також використовують інформаційний критерій Акайке

$$IKA = N \ln\left(\sum_{k=1}^{N} e^2(k)\right) + 2n$$

та критерій Байєса-Шварца

$$KBS = N\ln\left(\sum_{k=1}^{N} e^2(k)\right) + n\ln(N),$$

де $n = p+q+1$ – число параметрів моделі, які оцінюються за допомогою статистичних даних (p – число параметрів авторегресійної частини моделі; q – число параметрів ковзного середнього; 1 з'являється тоді, коли оцінюється зміщення, тобто a_0).

У правій частині виразів знаходиться сума квадратів похибок, а тому за цими критеріями вибирають ту модель, для якої критерії приймають найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується n), але разом з тим зменшується сума квадратів похибок і критерій в цілому зменшується. Якщо регресор не покращує модель, то критерій збільшується. Необхідно також відзначити, що асимптотичні властивості для довгих вибірок кращі у критерії Байєса-Шварца.

е) Окрім згаданих параметрів, для визначення адекватності моделі використовують коефіцієнт множинної детермінації R^2, F-статистику Фішера та статистику Дарбіна-Уотсона (для перевірки корельованості похибок).

Коректне використання методики Бокса-Дженкінса забезпечує побудову адекватної математичної моделі процесу, якщо експериментальні дані відповідають вимогам представництва та інформативності. Перша вимога означає, що вибірка даних повинна охоплювати досить довгий проміжок часу, щоб перекривати ті режими функціонування процесу, які цікаві для дослідника. Вимога інформативності означає, що вибірка повинна вміщувати в собі кількість інформації, достатню для оцінювання коефіцієнтів моделі. Наприклад, якщо моделюється процес другого порядку, то вибірка повинна забезпечувати обчислення першої та другої похідної. Умову інформативності ще називають умовою достатнього збудження процесу.

Ідентифікація моделей ARIMA-АРПКС зсувних процесів Південного берега Криму

Ключове завдання аналізу зсувних процесів Південного берега Криму полягає в попередній оцінці і подальшому якісному прогнозі зсувної активності та об'єму коштів для попередження або подолання катастрофічних наслідків цих процесів. Заздалегідь сплановане укріплення зсувонебезпечних ділянок дозволяє раціонально розподілити ресурси, які виділяються на протизсувні роботи в регіоні, не зачіпаючи при цьому інтересів населення та відпочивальників на Південному березі Криму. Подібна політика приводить до виключення катастрофічних ситуацій, які загрожують життю та господарській діяльності, а також до підвищення рейтингу регіону, який використовується з рекреаційною, оздоровчою та туристичною метою, завдяки повній відповідності потребам споживачів цих послуг.

Організація даних

На Південному березі Криму відповідні структури ведуть спостереження за зсувними процесами та накопичують дані, які їх описують. Загальноприйнято досліджувати відсоток зсувів, що активізувалися впродовж року, проте в моделях, що будуть побудовані, є сенс враховувати не відсоток, а загальну кількість активних зсувів. Це пов'язано з тим, що кожного року появляються нові зсуви і вносяться до кадастру, а старі (чи активні, чи стабілізовані) також впливають на загальну кількість, тому відносна кількість зсувів, врахована у відсотках, не зовсім об'єктивно описує ці процеси. Розглянемо дані з кількості активних зсувів на Південному березі Криму за кожен рік за

період з 1962 по 2005 роки. Дані були представлені в таблиці Excel і експортовані в STATISTICA. Таким чином, необхідно провести аналіз однієї змінної з 44 спостереженнями.

Попередній аналіз

Спочатку побудуємо початковий ряд графічним методом [3], а також гістограму розподілу накопичених даних (рис. 1).

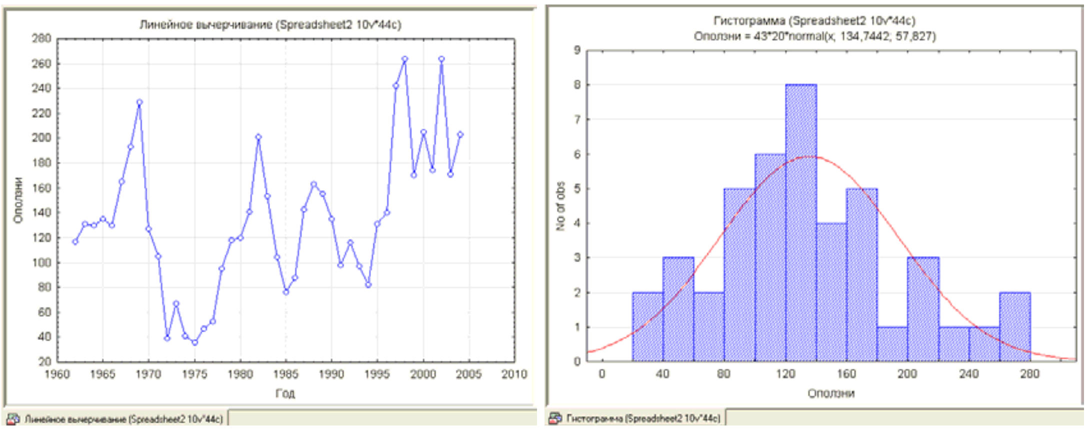


Рисунок 1 – Лінійний графік і гістограма кількості активних зсувів

З графіка змінної чітко бачимо, що існує сезонна залежність кількості активних зсувів. Гістограма ілюструє той факт, що розподіл ряду не є нормальним, але близький до нього, і має сенс спробувати встановити залежності для цього ряду.

Аналіз даних

Знаходження залежності в представлених даних являє собою завдання: розбити початковий ряд на 2 складові – детерміновану функцію і чисто випадкову складову. Випадкова складова повинна бути рядом Гауса з незалежними прирощеннями.

Для встановлення характеру невипадкової складової побудуємо автокореляційну та часткову автокореляційну функції початкових даних (рис. 2).

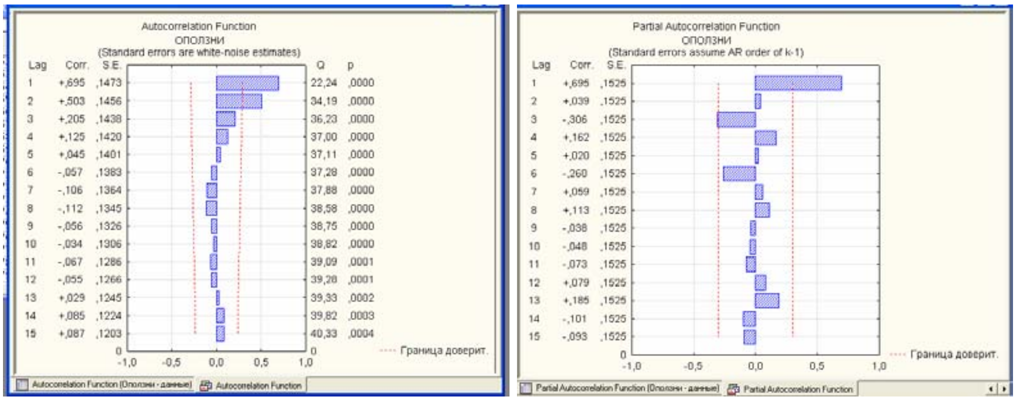


Рисунок 2 – Автокореляційна функція (АКФ) і часткова автокореляційна функція (ЧАКФ)

Автокореляційна функція має тенденцію до згасання, що говорить про стаціонарність процесу. Стаціонарні ряди мають постійні за часом середнє, дисперсію й автокореляції (тобто сезонні залежності видаляються за допомогою різниць [4]). Для моделі АРПКС необхідно, щоб ряд був стаціонарним, це означає, що його середнє постійно, а вибіркова дисперсія й автокореляція не міняються з часом.

За виглядом автокореляційної функції і часткової автокореляційної функції можна припустити, що ряд описується моделлю авторегресії 1-го порядку.

Розглянемо декілька моделей аналізу часових рядів АРПКС (авторегресії та проінтегрованого ковзного середнього):

– авторегресія 1-го порядку АР(1) або АРПКС(1)

$$Y_i=b_0 + b_1*Y_{i-1};$$

де Y_i , Y_{i-1} – кількість активних зсувів у відповідному році;

– авторегресія 4-го порядку АР(4) або АРПКС(4)

$$Y_i = b_0 + b_1*Y_{i-1} + b_2*Y_{i-2} + b_3*Y_{i-3} + b_4*Y_{i-4};$$

– авторегресія 1-го порядку і ковзного середнього 1-го порядку АРПКС(1,1)

$$Y_i = b_0 + b_1*Y_{i-1} + b_2*a_{i-1};$$

де a_{i-1} – білий шум;

– авторегресія 2-го порядку і ковзного середнього 1-го порядку АРПКС(2,1)

$$Y_i=b_0 + b_1*Y_{i-1} + b_2*Y_{i-2} + b_3*a_{i-1};$$

– авторегресія 1-го порядку і ковзного середнього 2-го порядку АРПКС(1,2)

$$Y_i=b_0 + b_1*Y_{i-1} + b_2*a_{i-1} + b_3*a_{i-2}.$$

Аналіз якості моделей проведемо за такими критеріями[3]:

1) коефіцієнт детермінації;
2) середній квадрат залишків;
3) статистика Дарбіна-Уотсона.

Аналіз якості прогнозу проведемо за такими критеріями:

1) середнє квадратичне відхилення – СКВ:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2};$$

2) середня абсолютна похибка – САВ:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i - \tilde{y}_i |,$$

де y_i – значення періоду, $\tilde{y}_i$ – прогнозна величина для i = 1, 2, …, n;

3) середня абсолютна відсоткова похибка – САВП:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \tilde{y}_i|}{|y_i|}\cdot 100\%;$$

4) коефіцієнт нерівності Тейла U: $U = \dfrac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i)^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i)^2}}$.

Результати оцінювання якості моделі та якості прогнозу для запропонованих моделей АРПКС наведені в табл. 1.

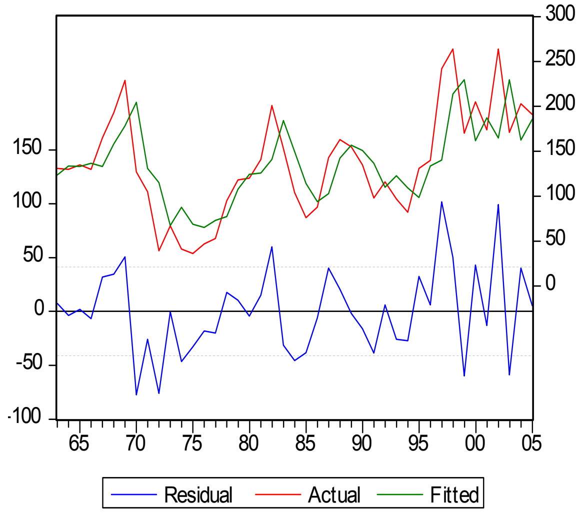
Таблиця 1 – Аналіз якості моделі та якості прогнозу

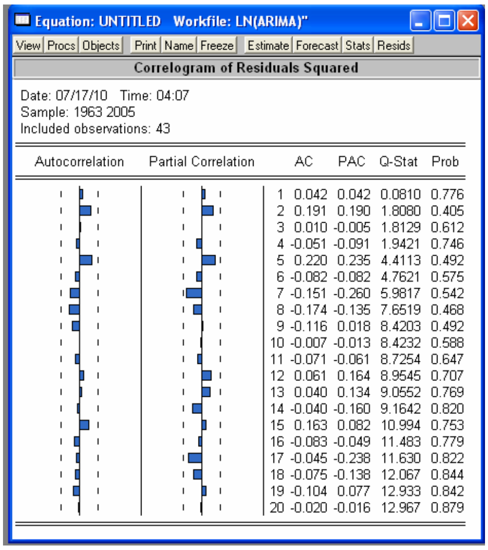
Моделі АРПКС	Якість моделі			Якість прогнозу			
	R2	Σe2 / n	DW	СКВ	САВ	САВП	К-т Тейла U
АРПКС(1)	0,501	1625	2,020	40,32	31,39	28,8%	0,139
АРПКС(4)	0,555	1557	1,985	40,04	32,82	29,4%	0,137
АРПКС(1,1)	0,512	1662	1,983	40,75	31,86	29,3%	0,140
АРПКС(2,1)	0,523	1625	2,022	40,31	31,42	29,2%	0,138
АРПКС(1,2)	0,546	1583	2,053	39,79	32,07	28,5%	0,136
АРПКС_L(1,2)	0,595	1528	1,960	48,71	32,01	26,0%	0,135

З наведених оцінок неможливо обрати модель, яка би значно відрізнялась від інших за однією чи декількома характеристиками. Значення, наведені в стовпчиках, май-

же співпадають, або відрізняються зовсім мало, що говорить про однаковість моделей. Очевидно, що всі моделі майже еквівалентні, тобто додавання складності в модель не привело до значного поліпшення оцінок моделі або оцінок прогнозу. Таким чином, можна зробити висновок: з наведених моделей найкраща та модель, яка є більш простою, тобто авторегресія 1-го порядку – АРПКС(1).

Всі розглянуті моделі для даного процесу назвемо умовно адекватними тому, що статистики, які їх характеризують, ледве досягають прийнятного значення, допустимого для подальшого прогнозування. Отже, отриману модель авторегресії 1-го порядку слід використовувати для прогнозування, але краще в сукупності з іншими моделями, які побудовано методами, що відрізняються від аналізу часових рядів АРПКС. Прогнози за різними моделями можна узагальнити і використовувати як прогноз, побудований за однією найбільш адекватною моделлю. Наведемо на рис. 3 результати прогнозування за допомогою моделі АРПКС(1): прогнозне значення (Fitted), величину, що спостерігалася, (Actual) та залишки (Residual), а також автокореляційну функцію (АКФ) і часткову авторегресійну функцію (ЧАКФ) залишків даної моделі.





Equation: UNTITLED Workfile: LN(ARIMA)"

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals Squared

Date: 07/17/10 Time: 04:07
Sample: 1963 2005
Included observations: 43

	AC	PAC	Q-Stat	Prob
1	0.042	0.042	0.0810	0.776
2	0.191	0.190	1.8080	0.405
3	0.010	-0.005	1.8129	0.612
4	-0.051	-0.091	1.9421	0.746
5	0.220	0.235	4.4113	0.492
6	-0.082	-0.082	4.7621	0.575
7	-0.151	-0.260	5.9817	0.542
8	-0.174	-0.135	7.6519	0.468
9	-0.116	0.018	8.4203	0.492
10	-0.007	-0.013	8.4232	0.588
11	-0.071	-0.061	8.7254	0.647
12	0.061	0.164	8.9545	0.707
13	0.040	0.134	9.0552	0.769
14	-0.040	-0.160	9.1642	0.820
15	0.163	0.082	10.994	0.753
16	-0.083	-0.049	11.483	0.779
17	-0.045	-0.238	11.630	0.822
18	-0.075	-0.138	12.067	0.844
19	-0.104	0.077	12.933	0.842
20	-0.020	-0.016	12.967	0.879

Рисунок 3 – Побудова авторегресійної моделі та АКФ і ЧАКФ

Пошук інших авторегресійних моделей

Перетворимо дані таким чином, щоб зменшити амплітуду коливання і видалити уклін тренда. Попереднім перетворенням *логарифмування* «погасимо» слабо виражену мультиплікативну природу початкового ряду, після цього віднімемо тренд. Ці дії автоматизовані в системі STATISTICA [3], і доступні в меню інші перетворення і графіки (рис. 4).

Візуальний аналіз результатів показує, що коливання випадкової величини відбувається навколо «нуля» та не перевищує одиниці за амплітудою, що свідчить про стаціонарність процесу.

Підберемо параметри моделі, вибравши модель АРПКС(2,3), тобто $p = 2$ – авторегресія 2-го порядку і $q = 3$ – ковзне середнє 3-го порядку:

```
Variable: ОПОЛЗНИ : ln(x); x-4,477-,015*t
 Transformations:
                Model: (2,0,3)
 No. of obs.:  43     Initial SS= 9,0324 Final SS= 2,9724(32,91%)  MS= ,07822
  Parameters (p/Ps-Autoregressive, q/Qs-Moving aver.); highlight: p<.05
             p(1)   p(2)   q(1)   q(2)   q(3)
   Оценка:   1,3994 -,5044 ,60215 -,4332 ,83025
   Std.Err.: ,16513 ,16454 ,08510 ,13395 ,08008
```

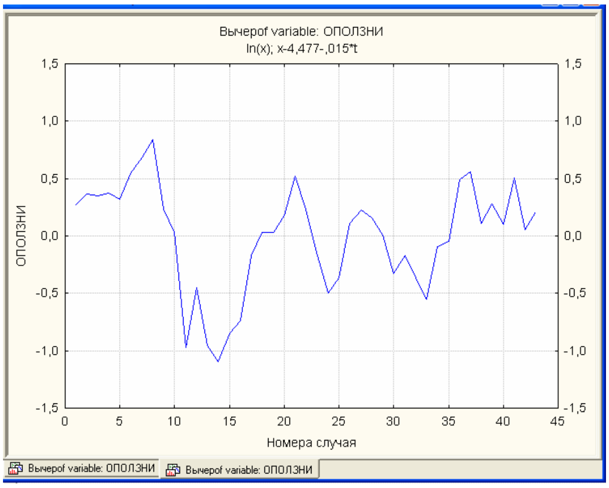


Рисунок 4 – Візуальний аналіз результатів

Спроможні оцінки параметрів підсвічуються сірим кольором.

Workbook3* - Вход: (Оползни - данные)

Вход: (Оползни - данные)
Transformations: none
Модель(2,0,3) MS Residual=,07822

Paramet.	Парамет	Asympt. Std.Err.	Asympt. t(38)	p	Нижний 95% Conf	Верхни 95% Conf
p(1)	1,399388	0,165130	8,47444	0,000000	1,065099	1,733678
p(2)	-0,504449	0,164535	-3,06590	0,003983	-0,837533	-0,171365
q(1)	0,602154	0,085097	7,07612	0,000000	0,429885	0,774423
q(2)	-0,433193	0,133946	-3,23408	0,002527	-0,704353	-0,162033
q(3)	0,830250	0,080078	10,36799	0,000000	0,668140	0,992360

Вход: (Оползни - данные)

Рисунок 5 – Оцінки параметрів

На рис. 5 наведено оцінки параметрів – 1 стовпчик, стандартні відхилення (2 ст.), t – статистика Стьюдента (3 ст.), ймовірності відхилення цих параметрів (4 стовпчик «р», в ньому всі значення менші за 0,005, тобто 0,5%) та нижні й верхні границі оцінок при рівні довіри 95% (5 ст.). Середній квадрат залишків становить 0,07822.

Для перевірки адекватності моделі скористаємося візуальними методами, представленими в системі STATISTICA. Як вже було сказано, випадкова складова – залишки – повинні бути нормально розподілені (рис. 6).

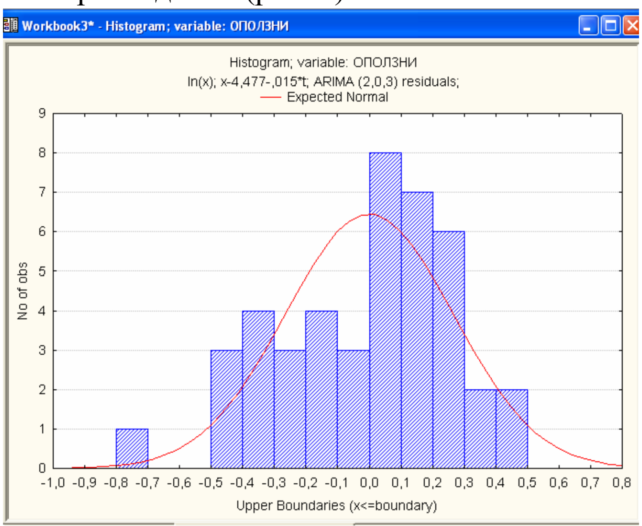


Рисунок 6 – Гістограма залишків моделі АРПКС(2; 0; 3)

Внаслідок того, що прирощення випадкової складової повинні бути незалежні, автокореляції і часткові автокореляції залишків не повинні виходити за допустимі інтервали (рис. 7).

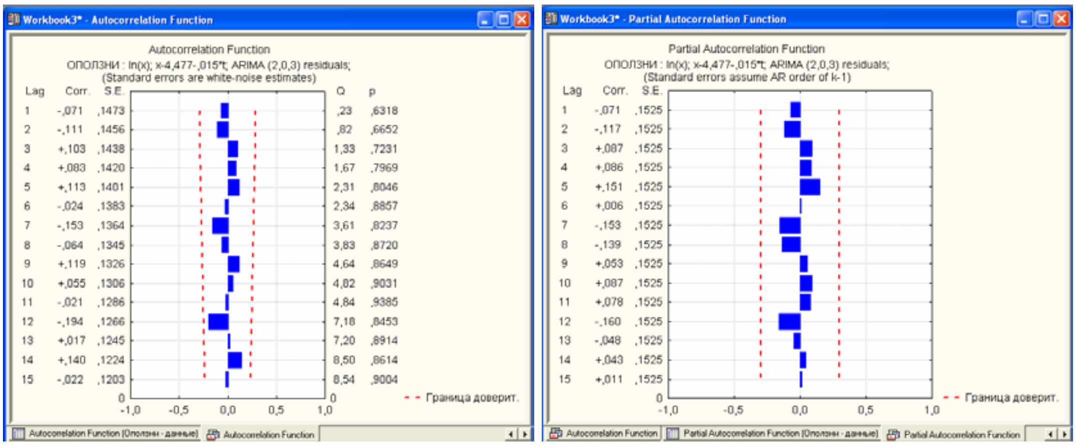


Рисунок 7 – Автокореляційна і часткова автокореляційна функції для залишків АРПКС моделі

Як видно, автокореляції і часткові автокореляції (рис. 7) цілком лежать в допустимих інтервалах. (У правильно підібраній моделі залишки будуть схожі на білий шум: у них не буде періодичних коливань, систематичного зсуву, між ними не буде сильних кореляцій).

Додатковим методом аналізу адекватності моделі є графік розташування залишків на нормальному імовірнісному папері (рис. 8).

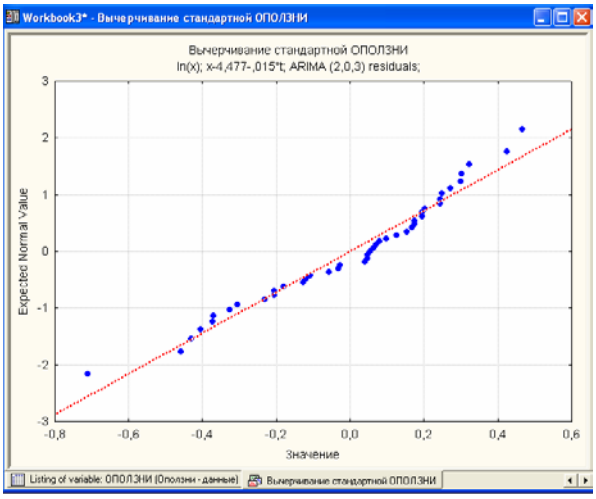


Рисунок 8 – Нормальний графік для залишків моделі АРПКС (2; 0; 3)

Точки на графіку лежать близько до прямої, що добре характеризує модель.

Отже, модель достатньо достовірна. Подивимося, наскільки точно вона передбачає дійсні спостереження. Для цього побудуємо прогноз за першими 75% спостережень на тих, що залишилися 25%. 75% спостережень складають 33 роки, 25% – 11 років.

Побудувавши графік прогнозу з 34 по 44 рік, бачимо, що прогноз добре передбачає дійсні спостереження, особливо в першому періоді прогнозу. За прогнозом з 44 року на 25% вперед інтуїтивно видно, що модель достатньо об’єктивна, тим більше, що за даними останніх п’яти років спостерігалося значне зниження активізації зсувних процесів.

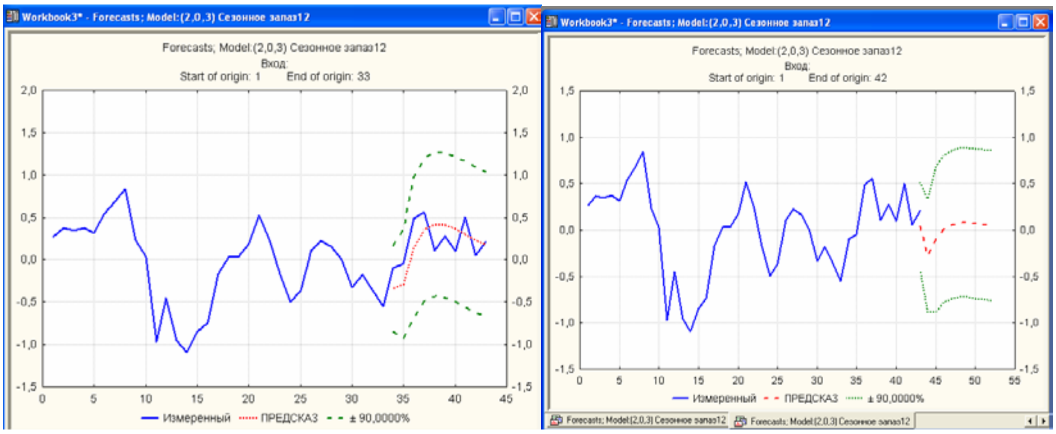


Рисунок 9 – Графік прогнозу на 10 років назад і на 10 років вперед за допомогою моделі АРПКС(2, 3)

Порівняння і аналіз різних моделей

Введемо позначення АРПКС_Л – авторегресія з проінтегрованим ковзним середнім логарифмована. Побудуємо і оцінимо моделі АРПКС_Л різних порядків. Результати оцінювання запишемо в узагальнюючу таблицю (табл. 2).

Таблиця 2 – Моделі АРПКС прологарифмовані із видаленим трендом

	Якість моделі			Якість прогнозу			
	R2	Σe2/n	DW	RMSE	MAE	MAPE	U
АРПКС_Л(1,1)	0,547	0,095	2,030	0,309	0,244	146,0%	0,387
АРПКС_Л(1,2)	0,595	0,086	2,017	0,361	0,283	135%	0,568
АРПКС_Л(1,3)	0,590	0,088	1,997	0,354	0,282	141%	0,550
АРПКС_Л(1,4)	0,586	0,089	1,960	0,359	0,285	143%	0,554
АРПКС_Л(2,3)	0,598	0,087	2,012	0,311	0,245	134%	0,487

Спроби провести аналіз з іншими параметрами не дали такої високої адекватності для моделі. Прогнозування залишків для підвищення точності аналізу виявилося досить складним завданням, оскільки залишки розподілені практично рівномірно та їх автокореляції близькі до нуля. Виділити сезонну складову у ряді залишків не вдалося.

Такий прогноз можна вважати середньотерміновим тому, що термін прогнозування становить досить тривалий відрізок часу – від року до 10. Це пов’язано зі специфікою спостережень за зсувними процесами (у зв’язку з обмеженим фінансуванням дані збираються раз на рік). Отже, модель АРПКС дає короткотермінові прогнози, тобто дає розрахунки на один крок вперед, але прогнози можна будувати і на триваліші терміни при наявності попередніх даних.

Проте впродовж року можуть відбуватися значні зміни факторів, що впливають на зсувні процеси Південного берега Криму, тому є сенс в побудові нової моделі, яка може вчасно реагувати на різні збурення факторів, тобто модель надаватиме короткотерміновий прогноз, який включатиме строк від одного тижня до двох – трьох місяців. Цим вимогам цілком відповідає мережа довіри Байєса [5].

Висновки

У статті розглянуто аналіз часових рядів за допомогою АРПКС моделей зсувних процесів на Південному березі Криму. Наведена методика Бокса-Дженкінса для побудови математичної моделі процесу у вигляді різницевого рівняння. Розглянуто детально етапи побудови моделі.

Проведено ідентифікацію моделей ARIMA-АРПКС зсувних процесів Південного берега Криму і визначено, що при ускладненні моделі додатковими параметрами оцінки адекватності моделі й адекватності прогнозу не покращуються.

Розрахунки оцінок моделей АРПКС(1), АРПКС(1,1), АРПКС(1,2), АРПКС(1,3), АРПКС(1,4) підтверджують результати попереднього аналізу. З таблиці видно, що моделі мають майже однакові статистичні характеристики у порівнянні одна з одною. Оцінки коефіцієнтів моделі суттєво відрізняються від нуля на рівні 1%, а корені характеристичного рівняння знаходяться всередині кола одиничного радіуса. Значення *Q*-статистики свідчать про те, що автокореляція між похибками є статистично несуттєвою, тобто нуль-гіпотеза $Q = 0$ підтверджується. Побудовано графіки залишків автокореляційної та часткової автокореляційної функцій (ЧАКФ).

Перед оцінюванням наступної моделі проведена попередня обробка даних – логарифмування та видалення тренду. Іншим підходом до попередньої обробки даних може бути використання квадратного кореня значень або іншої процедури.

У вищенаведених моделях сезонний ефект не був описаний за допомогою лагу. Крім такого підходу до моделювання сезонного ефекту існують також і інші, які можуть дати точніше описання процесу. Наприклад, існують моделі з мультиплікативними сезонними компонентами

$$(1 - a_1 L)\, y(k) = (1 + \beta_1 L)(1 + \beta_4 L^4)\, \varepsilon(k),$$

де $\beta_4 \varepsilon(k-4)$ входить в модель у мультиплікативній формі. Така модель може мати кращі показники прогнозування. Наступним кроком дослідження процесу може бути тестування часового ряду на наявність гетероскедастичності, тобто чи є дисперсія ряду змінною величиною.

Література

1. Системний аналіз екологічно небезпечних процесів різної природи / Ю.М. Сєлін // Системні дослідження та інформаційні технології. – 2007. – № 2. – С. 22-32.
2. Бідюк П.І. Аналіз часових рядів / Бідюк П.І. – Київ : ННК ІПСА, 2006. – 188 с.
3. Таран В.М. Методика оцінювання регресійних моделей, побудованих за даними спостережень, що описують зсувні процеси Південного берега Криму / В.М. Таран // Матеріали XII Міжнародної науково-технічної конференції «Системний аналіз та інформаційні технології». Київ 2010, САІТ – 2010. – 159 с.
4. Таран В.М. Використання інтелектуальних систем при прогнозуванні зсувних процесів Південного берегу Криму / В.М. Таран // Искусственный интеллект. – 2006. – № 3. – С. 441-449.
5. Таран В.М. Моделювання зсувних процесів Південного берега Криму в умовах невизначеності / В.М. Таран // Нові технології. – 2007. – № 1 – 2 (15 – 16). – С. 259-265
6. Боровиков В.П. Прогнозирование в системе STATISTICA в среде WINDOWS / В.П. Боровиков, Г.И. Ивченко. – Москва : Финансы и статистика, 1999. – 384 с.
7. Прогноз числа авиапассажиров методами временных рядов в системе STATISTICA [Электронный ресурс]. – http:// www.statsoft.ru/ statportal/ tabID__71/ MId__330/ ModeID__0/ PageID__135/ DesktopDefault.aspx
8. Лук'яненко І.Г. Економетрика / І.Г. Лук'яненко, Л.І. Краснікова [підручник]. – К. : Знання, 1998. – 494 с.
9. Круцик М.Д. Захист гірських автомобільних доріг від зсувів / Круцик М.Д. – Коломия, 2003. – 425 с.
10. Козаченко Ю.В., Пашко А.О. Моделювання випадкових процесів / Ю.В. Козаченко, А.О. Пашко. – Київ : Київський університет, 1999. – 224 с.

В.Н. Таран, А.О. Пашко

Оценка прогнозов оползневых процессов Южного берега Крыма с помощью анализа временных рядов (АРПСС)

В статье исследованы общие закономерности построения авторегрессионных моделей оползневых процессов на примере Южного берега Крыма. Проведена идентификация АРПСС-моделей, выполнены расчеты оценок параметров этих моделей. Выбраны критерии оценки моделей и оценки прогнозов, по которым выделяется модель среди остальных, а также построен прогноз на 10 лет.

Estimation of Prognoses of Landslide Processes of the Southern Coast of Crimea by Means of Analysis of Temporal Rows (ARIMA)

General conformities with a law of construction of autoregressive models of landslide processes are investigated, on the example of the Southern Coast of Crimea. Authentication of ARIMA-models is conducted, the calculations of estimations of parameters of these models are executed. The criteria of estimation of models and estimation of prognoses on which a model is distinguished among the others are chosen, and also a prognosis is built for 10 years.

Стаття надійшла до редакції 19.07.2010.